

Internet Data Collection

FINAL REPORT

This evaluation reports the results of research and analysis undertaken by the U.S. Census Bureau. It is part of a broad program, the Census 2000 Testing, Experimentation, and Evaluation (TXE) Program, designed to assess Census 2000 and to inform 2010 Census planning. Findings from the Census 2000 TXE Program reports are integrated into topic reports that provide context and background for broader interpretation of results.

Erin Whitworth

Decennial Statistical
Studies Division

CONTENTS

Executive Summary	iii
1. BACKGROUND	1
1.1 Had the Internet been used in a census before?	1
1.2 What were some of the development challenges to overcome?	1
1.3 Why did the Census Bureau not advertise this response mode?	2
1.4 What was the processing flow?	2
2. METHODOLOGY	3
2.1 What is the purpose of this evaluation?	3
2.2 What were the data files used for this evaluation?	3
3. LIMITATIONS	3
4. RESULTS	4
4.1 What is the profile of eligibility and initial response?	4
4.2 What is the profile of Internet usage?	6
4.3 How many multiple responses did we receive?	8
4.4 What were the demographics for PSA-selected households responding <i>only</i> on the Internet?	10
5. RECOMMENDATIONS	17
REFERENCES	18
APPENDIX A	19
APPENDIX B	22

LIST OF FIGURES AND TABLES

Figure 1. Time elapsed in filling out the form	6
Figure 2. Percentage of Internet forms received daily between March 3 and April 18	8
Figure 3. Distribution of multiple Internet returns	9
Table 1. Status of Internet submissions in processing steps	7
Table 2. Type of Enumeration Area	11
Table 3. Tenure	11
Table 4. Reported household size	13
Table 5. Sex	14
Table 6. Age	15
Table 7. Race	16
Table 8. Hispanic origin	17

Executive Summary

Census 2000 was the first U.S. census ever to include an Internet on-line reporting capability. Though it was met with many challenges, the Internet data collection was an operational success. Given the low volume of responses, it proved to be secure, and there were no hardware or software failures and no known security breaches. However, it did not fully reach its potential. This is probably because of a conscious decision by the Census Bureau not to advertise this response mode. The Internet data collection system could have handled tens of millions of forms. Instead,

- We received 89,123 initial on-line requests for an Internet census form.
- About 16.7 percent of these initial requests were invalid (mostly requests for the long form).
- There were 63,053 households representing 169,257 persons that were counted using just the Internet census form.

Less than four percent of Internet submissions had other returns from different modes. About two-thirds of the households returning a second Internet submission were submitting duplicate data. This probably occurred because the respondent simply hit the “Submit” button more than once without changing any of the data.

The Internet form collected data for six persons, with a continuation roster similar to the short-form paper questionnaire. This restriction of a six-person form required followup for large households, but did not greatly affect Coverage Edit Followup operations. Less than two percent of the Internet households reported household size to be larger than six.

Internet households had approximately the same reported household size as reported for all response modes. There were some differences between the demographic makeup for the Internet household members when compared to the overall population. This is not surprising when one considers the typical Internet user. Demographic highlights include:

- There were slightly more males (52.0 percent) than females (48.0 percent).
- There were a higher percentage of white, and non-Hispanic household members than for all response modes.
- There were a higher percentage of persons between the ages of 25-54 responding on the Internet as compared to all response modes.

The Internet is here to stay. The exact form and function of Census Internet options, however, is largely undeveloped. The software and hardware developed for this program could have handled **tens of millions** of records instead of the tens of thousands it did handle. It is our recommendation that future research focus not only on **how** to securely implement the form itself, but also focus on how to promulgate the Internet form as a major response option. Future research should also focus on how to use it as a tool to increase data quality by implementing real-time data feedback and analysis.

1. BACKGROUND

This evaluation provides operational summaries on the use of the Internet as a response mode in Census 2000. The evaluation will guide the reader through the processing of Internet returns, namely:

- the process to access the on-line census short-form questionnaire,
- the process by which we processed the Internet short-form questionnaires, and
- the process to determine those forms actually included in the census as Internet returns.

The terms “form,” “on-line census questionnaire,” and “form submission” will be used interchangeably. We restrict our analysis to those forms received by April 18, the cutoff for Nonresponse Followup (NRFU) operations.

1.1 Had the Internet been used in a census before?

The Internet had never been used in a previous census or census dress rehearsal or test to collect data. The Census 2000 dress rehearsal had a plan to include Internet Data Collection, but the Census Bureau abandoned this plan due to security concerns. The Department of Commerce, in the fall of 1998, decided that the Census Bureau should provide an opportunity for respondents to fill out their questionnaires on the Internet. Unfortunately, we had no background data on what the possible outcomes would be for this type of operation in a decennial census setting.

Respondents receiving the short form were able to respond on the Internet, if they could provide their 22-digit Census ID. There was insufficient time to create a Spanish-language version of the form for Puerto Rico. Thus, respondents in Puerto Rico and other Spanish speakers could respond on the Internet in English only. To submit a response on the Internet, the respondent:

- Began at the Census 2000 home page, <http://www.2000.census.gov>,
- Clicked on the link to the Internet form, and he or she was prompted to enter the 22-digit ID printed on the paper form,
- Received an on-line form similar to the paper form, with the same questions (if the 22-digit ID was valid),
- Completed the questions on the form, and
- Submitted the form by clicking on the button at the end of the form.

1.2 What were some of the development challenges to overcome?

The short amount of time available to develop this response mode presented challenges for the Census Bureau. There was simply not enough time to effectively test a long-form version of the questionnaire. Therefore, we provided only a short-form Internet response option, in English only. Many have questioned the wisdom of this decision, but it was necessary from a development standpoint. We needed to make this mode compatible with a wide range of browsers, and thus we were limited to using mostly HTML programming, and could not use JavaScript, which is unstable in some environments. JavaScript would have enabled us to have

multiple screens for a form, allowing for development of a long-form instrument. Putting the long form on one page would cause the form to take an inordinate amount of time to load on the user's screen. Having multiple screens presented some potential security risks that the Census Bureau was unwilling to take, especially as there was not sufficient time to test alternate methods.

We decided not to include real-time analysis, feedback, and editing on the on-line short form. For example, if a respondent reported that there were three people living or staying in the house, apartment, or mobile home on April 1, but then provided data for only two people, we would still accept the form. Another type of feedback would have helped identify discrepancies in the date of birth and age as of April 1. The absence of real-time edit checks was partly because our goal was to make the on-line form replicate, as much as possible, the paper form. We did, however, include some post-submission, preprocessing edits. For example, if respondents indicated that the month of birth was "DE," we assume that it corresponded to the month, "December." We then translated it to a numerical equivalent, namely, "12." Given more time for development, research, and experimentation, we would have incorporated some real-time analysis and feedback.

1.3 Why did the Census Bureau not advertise this response mode?

A major obstacle in this project was the lack of advertising by design and stated policy. The Census Operational Managers (COM) in July 1999 had some concerns related to public relations. They were also worried about possibly overburdening the Primary Selection Algorithm (PSA) process. Therefore, there was no indication anywhere on printed material that one could fill out the form on-line. The COM recommended that a press release should be the only means of advertising, yet this was never released due to disagreements over wording between the Department of Commerce and the Census Bureau. Unless someone happened to stumble across the link to the on-line form, or had some connection to the Census Bureau, he or she would not likely have known about this response mode. Even with the lack of advertising and promotion, we did not downscale the development of this response mode. The system was designed to handle **tens of millions** of responses instead of the **tens of thousands** we did receive. Given the low response we did receive, it was secure, safe, reliable, and a viable response option.

1.4 What was the processing flow?

This evaluation will take the reader on the same journey that successful Internet submissions followed. Refer to Appendix A for a flowchart that shows the different steps we used to create our analysis universe. We will refer to sections of this flowchart throughout the evaluation. Refer to Appendix B for a copy of the on-line version of the Census form.

2. METHODOLOGY

2.1 What is the purpose of this evaluation?

The overall objective for this evaluation is to document participation in the Census 2000 Internet Questionnaire between March 3 and April 18, 2000. This includes the documentation of the eligible universes, the responding universes and their respective demographics, and logistical aspects regarding access to the questionnaire, possible fraud detection, and multiple response resolution.

2.2 What were the data files used for this evaluation?

The following files were used for this evaluation:

- Hit count records from the Systems Support Division (SSD) were used to provide us a rough estimate of the volume of traffic to the Web site.
- Daily server logs from SSD were used to determine the distribution of Census ID submission attempts, and on-line short-form questionnaire attempts.
- Data files from SSD containing respondent-reported data were used as the baseline file of submitted on-line census forms.
- Daily status reports from the Decennial Systems and Contracts Management Office (DSCMO) provide us a record of the distribution of successfully submitted forms from SSD to DSCMO.
- The Decennial Response File 2 (DRF2) was used in conjunction with the SSD files to determine whether or not a particular Master Address File ID (MAFID) made it to this processing step. The MAFID is a 12-digit number that uniquely identifies a housing unit.
- Input and output files from the Coverage Edit Followup (CEFU) operation were used in conjunction with the SSD data files to determine which housing units were associated with CEFU.
- The Hundred Percent Edited File with the reinstated cases (HCEF_D') was used in conjunction with the SSD data files to determine which on-line census forms were accepted for the census. We also used the HCEF_D' to provide demographic characteristics for the final universe.

3. LIMITATIONS

There are no major limitations in the analysis presented in this evaluation. However, one respondent started filling out the census form shortly before midnight on April 18, and finished shortly after midnight on April 19. For the purposes of the evaluation, we will consider this

response received before the April 18 cutoff for Nonresponse Followup. Thus, the analysis numbers will be *slightly* affected, but this is not a problem for this analysis. Please keep this in mind when reviewing the results section.

This evaluation also examines the number of hits to the Census Bureau's Web site. However, we have no accurate measure of how many of the initial visitors to the main page containing the questionnaire continued to try to fill out an Internet version of the census form.¹

4. RESULTS

4.1 What is the profile of eligibility and initial response?

4.1.1 How many households were eligible for this response mode?

Only short-form Mailback households were eligible for this program (see Step 1 of Appendix A). There were a total of 89,536,424 stateside households eligible for this program. Puerto Rico added an additional 1,094,593 households. This number represents the total number of households that could potentially return a short form (U.S. Census Bureau, 2002). While it is naive to assume that all of these households had Internet access at the time of the census, we consider this "eligibility universe" to be the upper bound.

The U.S. Department of Commerce's Economic and Statistics Administration found that the share of households with Internet access increased from 26.2 percent in December 1998 to 41.5 in August 2000. Further, they found that the share of individuals **using** the Internet increased from 32.7 percent to 44.4 percent in the same period (U.S. Department of Commerce, 2000).

It is safe to assume that some households could access the Internet in places other than their own homes, such as places of work, libraries, or community organizations. It follows that our "eligible universe" probably was in the ballpark of 35 million households, based on a 40-percent Internet access rate. None of these descriptions of our "eligible universe" mean very much when we consider the true eligible universe. This universe is the one that **had** some kind of Internet access at the time of the census and **knew** about the on-line version of the census form. Thus, the true universe is **much** smaller than any of these numbers, and we will probably never know the exact magnitude of this universe (see Step 2 of Appendix A).

4.1.2 How many hits did the Census 2000 home page receive?

Estimates of the total number of hits to the main site, <http://www.2000.census.gov>, range from 19 million to almost 24 million during the March 3 through April 18 time frame. These

¹ Web page hits are not an accurate measure of Web traffic volume. They can be used as a relative measure of one page's hits relative to another page's hits, or one server's hits relative to another server's. There is NO WAY to accurately measure Web traffic by page or image hits. They are a poor measure, but in this case, the only measure. Caching servers that speed up the Web hide hits and users from the serving source, so that the more volume a page or system witnesses, the less likely it is to register hits accurately.

estimates include hits to every piece of information on the Web site, including GIF (Graphics Interchange Format) and PDF (Portable Document Format) files. Therefore, the estimates give us only a **rough** idea of the volume of traffic to the on-line form. However, they give a general starting point to examine the flow of responses at this site. The actual number of visitors is most likely between 10 and 30 percent of the 19 to 24 million hits. However, as mentioned previously, there is really no way to know for certain (see Step 3 of Appendix A).

On the initial page, there is a link for the user to follow to see if the paper form can be filled out on the Internet. We have no accurate measure of how many of the initial visitors to the main page continued to try to fill out an Internet version of the census form.

4.1.3 How many initial requests were made for the on-line form?

Between March 3 and April 18, 2000, there were 89,123 Census ID submissions, of which 16.7 percent (14,926 submissions) were failures (see Step 4 of Appendix A). It is quite possible that many, if not most, of the submission failures were associated with a Census ID representing a long form. The long-form sampling rate was, on average, one in six, or 16.7 percent. Since we did not advertise the Internet response option, respondents would have also had no idea that long-form households were ineligible. It is entirely probable that respondents in the long-form households attempted to submit their Census ID to request a form, and were denied access (see Step 5 of Appendix A). Other reasons that the ID submission attempts failed are:

- an invalid ID
- invalid browser information,
- a security error, or
- an expired form (this is explained in the following paragraph).

Most invalid IDs could be from long-form households or from respondents miskeying the 22-digit IDs. If the respondent's browser was set to suppress certain information that the census form required, or if there was a hiccup in the connection, the error would be attributed to invalid browser information. Other errors were attributed to security issues. If the user's browser did not support encryption, we included a disclaimer box that the user could check in order to proceed without the required level of encryption. If this box was not checked, the submission was logged as a security failure. If the ID page was on the user's screen for more than five minutes, and the user then tried to request a form, the user would get a message requesting that he or she try again. For security purposes, we only accepted form requests within a five-minute window. Form requests taking longer than five minutes were considered "expired forms."

There were 74,197 successful ID number submission attempts (see Step 5 of Appendix A). The respondent then received an on-line census form. However, of the successful **ID number** submission attempts, there were only 71,333 **on-line census questionnaire** submission attempts. It is quite possible that some of the remaining 2,864 successful ID number submission attempts represented respondents who had incorrectly keyed in a correct ID. These same respondents could have subsequently keyed in their ID correctly, and then continued with the process (see Steps 6 and 7 of Appendix A).

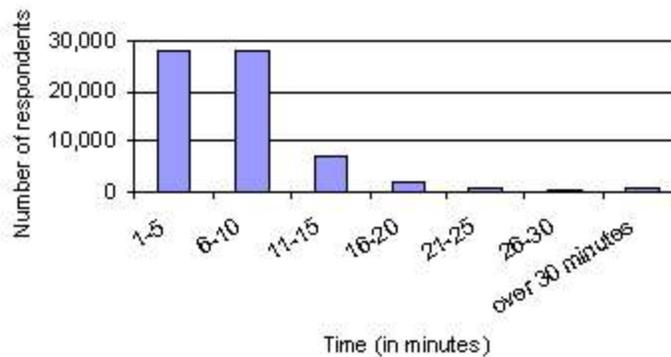
4.2 What is the profile of Internet usage?

4.2.1 How much time did it take to complete the questionnaire?

The amount of time respondents took to complete the form varied greatly. The best measure is the time from which the respondent received the form on-screen to the time the respondent submitted the form. This time does not take into account the possibility of the respondent taking a break during the response process, thereby increasing the measured elapsed time.

The great majority of respondents took less than ten minutes (the estimated time to complete the paper short form questionnaire); the average was 7.5 minutes. The times ranged from one minute to over 4.5 hours. It is quite possible that some respondents started filling out the form, and then left it up on their browser while they took a phone call, went to lunch, or took another unrelated break. Figure 1 shows the distribution of response times.

Figure 2. Time elapsed in filling out the form.



4.2.2 How many forms were submitted by April 18 and subsequently processed?

Of the 71,333 on-line census questionnaire submission attempts, SSD reported that they successfully received 69,765 successful form submissions by April 18 (see Step 8 of Appendix A). However, SSD received data for only 68,319 on-line census questionnaire submissions. There are no apparent explanations for this discrepancy (see Step 9 of Appendix A).

Before discussing the number of forms received by April 18 and subsequently processed, it is important to understand how many forms were received in total. There were 68,319 initial Internet forms collected between March 3, 2000, and April 18, 2000. Of these, 66,556 submissions were associated with unique MAFIDs (see Step 10 of Appendix A). The other 1,763 submissions were associated with MAFIDs that we had previously collected through the Internet response mode. Of these 66,556, SSD sent 66,163 forms to DSCMO for processing (see Step 11 of Appendix A).

Table 1 shows the distribution of Internet submissions that were sent to DSCMO for processing and those that were not. It also shows the distribution of forms accepted into the DRF2. A discussion of the data in this table follows the table (see Steps 11 and 13 of Appendix A).

Table 1. Status of Internet submissions in processing steps

Status on the DRF2	Totals	Processed by DSCMO	
		Yes	No
Totals	66,556	66,163	393
Included on the DRF2	66,064	65,683	381
Excluded from the DRF2	492	480	12

The careful reader might wonder why 480 processed Internet responses were not on the DRF2. This number can be found in the column labeled “Processed by DSCMO, YES,” in Table 1. A closer look at these particular returns gives us the answer. After obtaining address information from the Decennial Master Address File (DMAF) for these returns, we see that all 480 of these submissions were associated with MAFIDs that were “killed” during Census 2000 processing (U. S. Census Bureau, 2000d).

What is even more interesting, however, is that 393 of the unique submissions never even made it to DSCMO. The reader can see this number in the “Totals” row of Table 1. We originally thought that these had some sort of submission error, but an inspection of these cases shows that only ten of these had submission errors. Eight were completely blank forms, and two were special forms that were not meant for transmission anyway. Of the unique submissions that were not processed, 380 of them were submitted on March 13. Remember that this is the first day we started receiving forms from the Mailback universe. It is not unusual to see a spike on this day for any set of data related to the census, but we should have seen a similarly large number on the next day, which we did not. Refer to Figure 2 under section 4.2.3 for a view of this phenomenon. A careful examination of other variables for these nonprocessed submissions reveals no immediate anomalies or answers as to why they were not included in DSCMO processing.

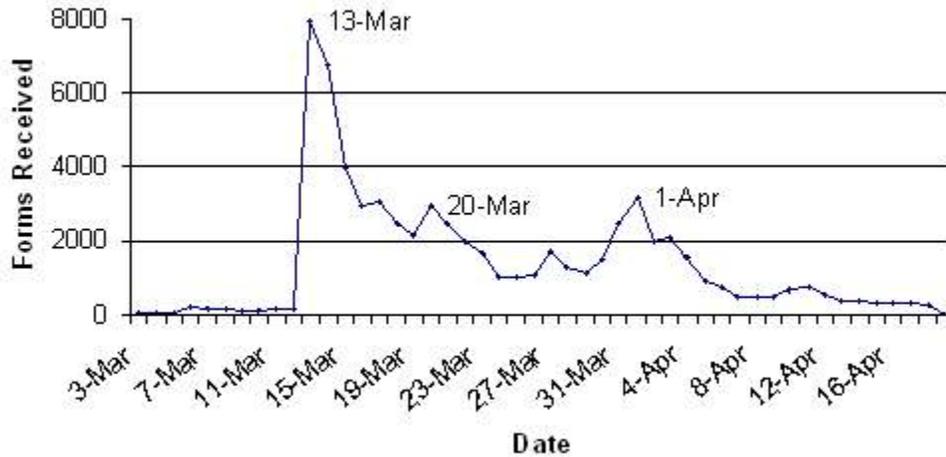
We will restrict our analysis to those Internet returns submitted and processed by April 18, and associated with MAFIDs appearing on the DRF2 file, thus resulting in 65,683 records.

4.2.3 When were the forms received?

The on-line version of the form was available from March 3 through April 18. Figure 2 shows the distribution of all forms received through the Internet. The first peak, on March 13, coincided with the mailing of the paper short-form questionnaires. Some households in Update/Leave areas already had their form before March 13, and subsequently submitted their form on-line. Further, it is likely that some households submitted their data by obtaining their 22-digit ID from the advance letter. The second peak coincides with the receipt of the reminder

postcard. Notice the peak on April 1; this was Census Day, and could also have been a perceived “deadline” by many respondents.

Figure 3. Percentage of Internet forms received between March 3 and April 18

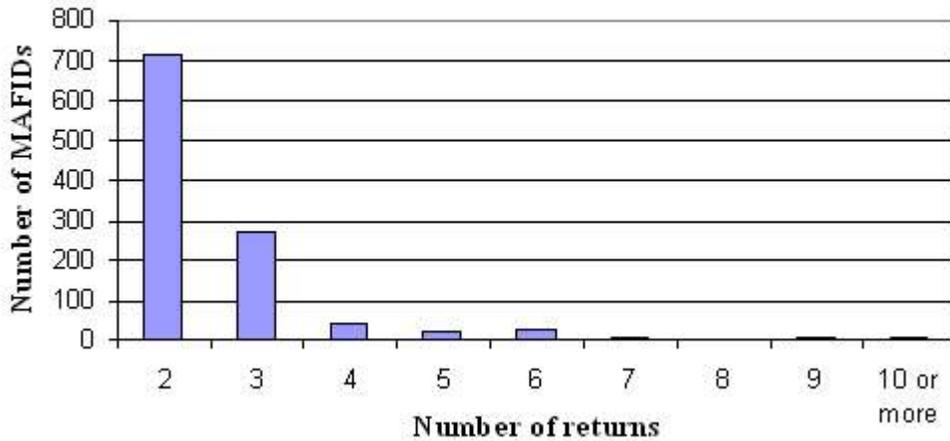


4.3 How many multiple responses did we receive?

4.3.1 How many multiple Internet forms did we receive (in the same session or multiple sessions)?

The vast majority of MAFIDs were associated with a single Internet response form. Of the 66,163 returns received by April 18, processed, **and** associated with unique MAFIDs, 65,073 submitted only one Internet form by April 18. The remaining 1,090 MAFIDs were associated with a total of 2,853 Internet form submissions by April 18. Figure 3 shows the distribution of multiple Internet responses. Keep in mind that these responses are not necessarily **duplicate** responses, but merely multiple responses associated with the same MAFIDs. However, some of the responses appear to be complete duplicates within a MAFID.

Figure 4. Distribution of multiple Internet returns



The first bar represents the number of MAFIDs associated with two Internet returns. A closer look at these multiple returns reveals some interesting things. Most of the data seem to be the same (looking at responses for basic questions). By examining the date and time of submission and the number of elapsed minutes from initiating the form until submitting the duplicate form, we have a better idea of the response process for these households. For example, several of these repeat submissions appear to be resubmissions of the same data. The respondent probably clicked on the “Submit” button several times. However, there appear to be others where the responses came from an entirely new response process, usually immediately after the first submission. A few were submitted on multiple days. It appears that only a few of the multiple response have data different from the first round of submissions. One respondent submitted three forms with information for two people, and then submitted three more forms with a third person added.

One of the most curious examples, however, is a MAFID associated with 17 (the highest in our dataset) different submissions. Many of these were on different days, and many with different data. Could there really be different people living in this same household that we missed because we did not process all of the submitted forms? An even closer look reveals that the answer is no.

4.3.2 How did the data vary for the multiple Internet returns?

It is somewhat comforting to note that of the 2,853 multiple Internet form submissions we received on the Internet by April 18, 62.9 percent (1,795 responses) were simply duplicate submissions of the entirely same set of response data. Of the remaining 37.1 percent of these multiple responses, only 13.7 (1058 responses) percent of them, (145 responses) responses, had reported differing household sizes.

Just above the “Submit” button, the form instructs respondents to verify their form for completeness and accuracy, tells them that they have only one opportunity to send the form, and instructs respondents to NOT resend the form once they receive a confirmation page. It is quite possible that we received some of these multiple responses from respondents who did not get a confirmation page. There really is not much we can do in a case where a respondent later decides to add or change information to the submitted form. Perhaps future instruments could give respondents the opportunity to add or change information through the use of coverage probes, such as those used in the Coverage Edit Followup.

4.3.3 How many Internet returns did we receive in addition to another response option?

This did not occur as often as one would expect. Of the 66,163 Internet returns received by April 18, processed, **and** associated with unique MAFIDs, only 2,373 had returns in addition to their Internet return. Over 95 percent of these MAFIDs had only one additional return. It appears, however, that over half of the additional returns associated with unique Internet returns were received after the April 18 cutoff for NRFU. Only 1,067, or 48.2 percent, of the MAFIDs had an additional return before the cutoff. For the most part, the additional returns were Mailback paper short-form questionnaires.

Regarding data quality of the submitted and processed Internet returns, we direct the reader to examine the study plan for the *Census 2000 Evaluation B.1, “Evaluation of the Analysis of the Imputation Process for 100 Percent Household Population Items”* (U.S. Census Bureau, 2001).

4.4 What were the demographics for PSA-selected households responding *only* on the Internet?

4.4.1 What is the geographic distribution of the households?

The PSA selected 65,243 of the 65,683 returns on the DRF2 for inclusion in the census (see Step 14 in Appendix A). As mentioned previously, this demographic analysis section will limit the scope to those Internet returns—

- received by April 18,
- identified on the DRF2 as Internet returns,
- selected by PSA for inclusion in the census, **and**
- that were the **only** return from a household.

As an example, the author returned her Internet form before April 18. It was identified on the DRF2 as an Internet return, it was selected by PSA for inclusion in the census, **but** she also returned a paper form. Thus, her Internet return is not in the scope of this analysis (see Steps 15, 16, and 17 of Appendix A). Therefore, we use a universe of 63,053 returns, representing the same number of households.

Table 2 shows the distribution of Internet households by Type of Enumeration (TEA) area as identified on the DRF2.

Table 2. Type of Enumeration Area

TEA	Frequency	Percent
Totals	63,053	100.0
Mailout/Mailback	56,842	90.2
Update/Leave	6,080	9.6
Military	30	0.1
Urban Update/Leave	101	0.2

4.4.2 What is the distribution of owners versus renters?

Table 3 shows that the percentage of renters versus owners is only slightly different from the percentages for all response modes. Notice the footnote explaining the reduction in the total by 145 vacant housing units. The careful reader might wonder why an Internet household is classified as a vacant housing unit. There are a few possible reasons for this. A respondent might have entered “0” for the household size. There might have been a count discrepancy between the number of reported persons and the data provided for those persons. These cases would have gone to Coverage Edit Followup, possibly resulting in a vacant housing unit because of residence rules. This small number of vacant housing units should not concern us.

Table 3. Tenure

Tenure	Internet		All Forms
	Frequency	Percent	Percent
Totals	62,908*	100.0	100.0
Owners	42,163	67.0	66.3
Renters	20,745	33.0	33.7

* The total is reduced by 145 vacant housing units.

We will now look at demographic characteristics associated with the person-level data obtained through the Internet form. We continue with our universe of 63,043 households.

4.4.3 How many forms were flagged for Coverage Edit Followup?

The Coverage Edit Followup (CEFU) program contacted households whose census return showed that there was a count discrepancy between the number of reported persons in the household and the number of people for whom data were provided. This program also contacted households that had more than six people listed on the form. Since there were no real-time edit checks, some of the Internet returns were also included in this program. Of the 63,053 forms described above, 1,709 of them went to Coverage Edit Followup. Approximately half, or 902 forms, were actually processed in the Coverage Edit Followup operation. Remember that our analysis universe is restricted to those households that returned **only** an Internet census form. We do not know exactly how many Internet households were contacted to resolve a count discrepancy. However, we do have the reported household size, which can give us an idea of how many Internet households might have been contacted to collect data for more than six persons on the form. The next section discusses the reported household size

4.4.4 What is the distribution of household size?

Most of the 63,053 PSA-selected Internet households answered the question, “How many people were living or staying in this house, apartment, or mobile home on April 1, 2000?” There were only 64 households that did not answer this question. Remember that the Internet form did not have built-in edit checks, which could have safeguarded against this problem. For example, with real-time edit checks, if a respondent entered character data in a numeric field, or left it blank, the Internet form would prompt the respondent to correct the response. Respondents **reported** a total of 174,017 persons. Table 4 shows the distribution of the reported household size. We compare this with the distribution of household size from all response modes.

Table 4. Reported household size

Household Size	Internet		All Forms
	Frequency	Percent	Percent
Totals	63,053	100.0	100.0
No response	64	0.1	—
0	467	0.7	9.0
1	12,825	20.3	23.4
2	19,565	31.0	29.6
3	11,504	18.2	15.1
4	11,003	17.5	13.0
5	4,811	7.6	6.0
6	1,660	2.6	2.3
More than 6	1,154	1.8	1.6

The table shows a higher percentage of households of size 2, 3, 4, and 5 for the Internet and a lower percentage of households of size 1. The average reported household size was 2.8, with responses ranging from 0 to 55. Compare this to a national average household size of 2.6 for all response modes. Remember that the 2.8 average is from the **reported** household size, which includes noise in the data, and is not the **final** average reported household size. Note that 1.8 percent of Internet households reported household sizes larger than six, which is the same percentage across all response modes. The census paper short form, *and* the Internet version, only had room for data for up to six persons, with a continuation name roster for an additional six. This small percentage of Internet respondents with larger households is an indication that we did not inordinately restrict data collection by limiting the size of the form on-line.

The 63,053 households represented a total of 174,017 reported persons, with a final person count of 169,257 persons selected for inclusion in the census. This second number is lower due to final editing and imputation processes.

4.4.5 What were the person characteristics of Internet household members?

This evaluation has a goal to report the characteristics of those **counted** through the Internet. Remember that we are restricting our analysis to those persons counted by April 18 on **only** an Internet form whose household was selected by PSA. Using the MAFIDs associated with these households, we can obtain demographic information for persons **in** these households on

person-level records on the HCEF_D'. We use this file as opposed to the DRF2 person-level file because the HCEF_D' has been edited and has the final coding and computation for age, sex, race, and Hispanic origin. The use of this file results in a total of 169,257 persons. The remainder of the analysis will focus only on these 169,257 persons. The following tables show the demographic characteristics of the Internet household members.

4.4.6 *What is the distribution by sex of all household members?*

Table 5 shows the distribution of sex for all household members **counted** through the Internet census form. We compare this with the distribution by sex for household members from all response modes.

Sex	Internet		All Forms
	Frequency	Percent	Percent
Totals	169,257	100.0	100.0
Male	87,932	52.0	49.0
Female	81,325	48.0	51.0

The table shows that Internet household members have slightly more males than females (52.0 percent males compared with 49.1 males for all response modes).

4.4.7 What is the distribution by age of all household members?

Table 6 shows the distribution by age of all household members counted through the Internet. We compare this with forms from all response modes.

Table 6. Age

Age	Internet		All Forms
	Frequency	Percent	Percent
Totals	169,257	100.0	100.0
Under 5 years	13,185	7.8	6.8
5 to 9 years	13,832	8.2	7.3
10 to 14 years	13,454	8.0	7.3
15 to 19 years	11,222	6.6	7.2
20 to 24 years	10,049	5.9	6.8
25 to 34 years	31,121	18.4	14.2
35 to 44 years	32,246	19.1	16.0
45 to 54 years	26,143	15.5	13.4
55 to 59 years	7,110	4.2	4.8
60 to 64 years	4,202	2.5	3.8
65 to 74 years	4,300	2.5	6.5
75 to 84 years	1,781	1.1	4.4
85 years and over	612	0.4	1.5

We see almost the same age distribution for Internet household members as we do for all response modes, except that there are slightly more in the 25-54 age groups, and fewer in the 55- and-over age groups. This is not surprising given the typical population of Internet users.

4.4.8 What is the distribution by race of all household members?

Table 7 shows the distribution of race of all household members counted through the Internet. We compare this with all response modes.

Race	Internet		All Forms
	Frequency	Percent	Percent
Totals	169,257	100.0	100.0
White	132,858	78.5	75.2
Black	15,185	9.0	12.3
American Indian or Alaska Native	1,038	0.6	0.9
Asian	8,693	5.1	3.6
Native Hawaiian or Other Pacific Islander	224	0.1	0.1
Some Other Race	6,587	3.9	5.5
Two or More Races	4,672	2.8	2.4

We see a higher percentage of whites and Asians in the Internet households and a lower percentage of blacks.

4.4.9 What is the distribution by Hispanic origin of all household members?

Table 8 shows the distribution of Hispanic origin for all household members counted through the Internet. Compare this with all response modes.

Table 8. Hispanic origin

Hispanic Origin	Internet		All Forms
	Frequency	Percent	Percent
Totals	169,257	100.0	100.0
Not Hispanic or Latino	153,957	91.0	86.3
Mexican	7,298	4.3	7.2
Puerto Rican	1,944	1.2	2.5
Cuban	659	0.4	0.4
Other Hispanic or Latino	5,399	3.2	3.6

We see a higher percentage of Internet household members who are not Hispanic compared with all response modes.

None of these demographic findings is very surprising, and the numbers make it even more obvious that the Internet is not a response mode that will improve coverage among traditionally undercounted groups. However, given the direction of technology, it appears that this response mode is here to stay. The exact form and function remains to be seen.

5. RECOMMENDATIONS

Obviously, the Internet is here to stay. The software and hardware developed for this program could have handled **tens of millions** of records instead of the **tens of thousands** it did handle. It is our recommendation that future research focus not necessarily on **how** to implement the form itself, but how to promulgate the Internet form as an option and convince the public that there is sufficient data security. Future research should also focus on how to use it as a tool to increase data quality by implementing real-time data feedback and analysis. The Internet option in Census 2000 was an operational success.

REFERENCES

- U.S. Census Bureau, 1997. *Decennial Census Internet Applications*, memo for the Census 2000 Internet Questionnaire Team, March 19, 1997.
- U.S. Census Bureau, 1997b. *Results From the Expert Review of the Defunct Decennial Census Electronic Short Form for the Web*, memorandum, November 18, 1997.
- U.S. Census Bureau, 1998. *Economic Data Collection via the Web: A Census Bureau Case Study*, Draft, August 6, 1998.
- U.S. Census Bureau, 2000a. Census 2000 Informational Memorandum, No. 5; *Clarification of Internet Programs for Census 2000*.
- U.S. Census Bureau, 2000b. *Program Master Plan: Census 2000 Internet Questionnaire Assistance and Data Collection Operations*, 2000.
- U.S. Census Bureau, 2000c. *Program Master Plan: Response Mode & Incentive Experiment*, 2000.
- U.S. Census Bureau, 2000d. *Specification of the Kill Universe on the Decennial Master Address File for Census 2000*, DSSD Census 2000 Procedures and Operations Memorandum Series #D-13, December 21, 2000.
- U.S. Census Bureau, 2001, *Evaluation of the Analysis of the Imputation Process for 100 Percent Household Population Items*, DSSD Census 2000 Procedures and Operations Memorandum Series #Y-1, October 1, 2001.
- U.S. Census Bureau, 2002, *Evaluation of the Census 2000 Mail Return Rates*, DSSD Census 2000 Procedures and Operations Memorandum Series, Draft.
- U.S. Department of Commerce, Economic and Statistics Administration, National Telecommunications and Information Division, *Falling Through the Net, Toward Digital Inclusion, A Report on Americans' Access to Technology Tools*, October 2000.